

Moment convergence of Z -estimators and Z -process method for change point problems

Ilia Negri and Yoichi Nishiyama*

*Department of Information Technology and Mathematical Methods, University of Bergamo
Viale Marconi, 5, 24044, Dalmine (BG), Italy*

ilia.negri@unibg.it

and

*The Institute of Statistical Mathematics
10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan*

nisiyama@ism.ac.jp

July 2, 2012

Abstract

The problem to establish not only the asymptotic distribution results for statistical estimators but also the moment convergence of the estimators has been recognized as an important issue in advanced theories of statistics. One of the main goals of this paper is to present a method to derive the moment convergence of Z -estimators as it has been done for M -estimators. Another goal of this paper is to develop a general, unified approach, based on some partial estimation functions which we call “ Z -process”, to the change point problems for ergodic models as well as some models where the Fisher information matrix is random and inhomogeneous in time. Applications to some diffusion process models and Cox’s regression model are also discussed.

1 Introduction

This paper is devoted to the study of two problems both based on the “ Z -methods”, in other words, some methods using the solutions to estimating equations. Let us first describe the outlines of the two themes, and next we shall list up some examples to which our results can be applied.

*Corresponding author.

1.1 Theme I: Moment convergence of Z -estimators

For an illustration, let us consider the simplest case of i.i.d. data. Let $(\mathcal{X}, \mathcal{A}, \mu)$ be a measure space, and let us be given a parametric family of probability densities $f(\cdot; \theta)$ with respect to μ , where $\theta \in \Theta \subset \mathbb{R}^d$. Let X_1, X_2, \dots be an i.i.d. sequence of \mathcal{X} -valued random variables from this parametric model. There are two ways to define the “maximum likelihood estimator (MLE)” in statistics. One way is to define it as the maximum point of the random function

$$\theta \mapsto \mathbb{M}_n(\theta) = \frac{1}{n} \sum_{k=1}^n \log f(X_k; \theta),$$

while the other is to do it as the solution to the estimating equation

$$\dot{\mathbb{M}}_n(\theta) = 0,$$

where $\dot{\mathbb{M}}_n(\theta)$ is the gradient vector of $\mathbb{M}_n(\theta)$. The former is a special case of “ M -estimators”, and the latter is that of “ Z -estimators”; see van der Vaart and Wellner (1996) for these terminologies. It may appear from the above introduction that M -estimation and Z -estimation can be regarded as “almost equivalent”. However, it is not always true as we will discuss below.

It is well known that the MLE $\hat{\theta}_n$ is asymptotically normal: it holds for any *bounded* continuous function $\psi : \mathbb{R}^d \rightarrow \mathbb{R}$ that

$$\lim_{n \rightarrow \infty} E[\psi(\sqrt{n}(\hat{\theta}_n - \theta_0))] = E[\psi(I(\theta_0)^{-1/2}Z)],$$

where $I(\theta_0)$ is the Fisher information matrix and Z is a standard Gaussian random vector. Furthermore, it is important for some advanced theories in statistics, including asymptotic expansions and model selections, to extend this kind of results for *bounded* continuous functions ψ to that for any continuous function ψ with *polynomial growth*, that is, any continuous function ψ for which there exist some constants $C = C_\psi > 0$ and $q = q_\psi > 0$ such that

$$|\psi(x)| \leq C(1 + \|x\|)^q, \quad \forall x \in \mathbb{R}^d. \quad (1)$$

See the discussion in Yoshida (2011) for the importance of this problem.

We observe that, when we have an asymptotic distribution result of an estimator, namely $R_n(\hat{\theta}_n - \theta_0) \rightarrow^d L(\theta_0)$ where R_n is a (possibly, random) diagonal matrix and the limit random vector $L(\theta_0)$ is not necessarily Gaussian, it is sufficient for the generalization to the case where ψ is a continuous function satisfying (1) to check that $\|R_n(\hat{\theta}_n - \theta_0)\|$ is *asymptotically L_p -bounded* for some $p > q$, that is,

$$\limsup_{n \rightarrow \infty} E[\|R_n(\hat{\theta}_n - \theta_0)\|^p] < \infty.$$

The study to provide some methods to obtain the moment convergence with polynomial order goes back to Ibragimov and Has'minskii (1981) who considered

the MLEs and the Bayes estimators (as some special cases of M -estimators) in the general framework of the locally asymptotically normal models. It should be emphasized that one of the important merits of Ibragimov and Has'minskii's program is that the theory, based on the likelihood, automatically yields also the asymptotic efficiency. In their main theorems, it was assumed that an *exponential type large deviation inequality* holds for the rescaled log-likelihood ratio random field. Kutoyants (1984, 1994, 1998, 2004) successfully applied this theory to different stochastic process models using some characteristics of the models under consideration such as small diffusion models, ergodic diffusion models and Poisson process models. However, developing a general theory to establish the large deviation inequality had been an open problem for many years. This problem was solved, and the results have been published in Yoshida (2011). The paper starts from pointing out that a *polynomial type large deviation inequality* is sufficient for the core part of Ibragimov and Has'minskii's (1981) program, and then the (polynomial type) large deviation inequality has been proved with a good generality. Uchida and Yoshida (2012) applied Yoshida's (2011) theory to establish the moment convergence of three kinds of adaptive M -estimators in ergodic diffusion process models including the one introduced by Kessler (1995). We mention that Nishiyama (2010) pointed out that the moment convergence problem for M -estimators can be solved also by using a maximal inequality instead of the large deviation inequalities, and that Kato (2011) took this type of approach to deal with some bootstrap M -estimators.

In this paper, we consider the problem of the moment convergence of Z -estimators. Since we have to assume that the random field (something like the log-likelihood) is differentiable, the framework of Z -estimation is more restrictive than that of M -estimation. On the other hand, our proof is a combination of arguments based only on usual Hölder's and Minkowskii's inequalities, and no large deviation type inequality appears in our treatment for Z -estimators.

Another difference between M - and Z -estimations is that in the latter theory the case where the rates of convergence are different over the components of θ can be treated easily. This is due to the fact that in the theory of Z -estimation the gradient vector $\dot{\mathbb{L}}_n(\theta)$ of a contrast function $\mathbb{L}_n(\theta)$, where $\mathbb{L}_n(\theta)$ is typically the log-likelihood function, can be pre-multiplied by a *matrix* R_n^{-2} to get a kind of law of large numbers, namely,

$$\dot{\mathbb{M}}_n(\theta) = R_n^{-2} \dot{\mathbb{L}}_n(\theta).$$

Typically, $R_n = \sqrt{n}I_d$ where I_d is the identity matrix, although in the approach to Z -estimation presented here the diagonal components of R_n may be different. Compare this with the framework of M -estimation where the (scalar valued) contrast function $\mathbb{L}_n(\theta)$ with no assumption of differentiability has to be multiplied by a scalar. Yoshida (2011) dealt with this point developing an iterative method.

Thinking of these differences, we may conclude that M -estimation and Z -estimation are not "almost equivalent" at least for the moment convergence problem.

1.2 Theme II: Z -process method for change point problems

Let us give an illustration by the example of independent data again. We introduce the partial sum process

$$\mathbb{M}_n(u, \theta) = \frac{1}{n} \sum_{k=1}^{[un]} \log f(X_k; \theta), \quad \forall u \in [0, 1],$$

and consider the gradient vectors $\dot{\mathbb{M}}_n(u, \theta)$ of $\mathbb{M}_n(u, \theta)$ with respect to θ . Let $\hat{\theta}_n$ be the MLE for the full data X_1, \dots, X_n as a special case of Z -estimators, that is, $\hat{\theta}_n$ is the solution to the estimating equation

$$\dot{\mathbb{M}}_n(1, \theta) = 0.$$

The fact that the random process

$$u \rightsquigarrow \sqrt{n} \dot{\mathbb{M}}_n(u, \theta_0) \quad \text{converges weakly to} \quad u \rightsquigarrow I(\theta_0)^{1/2} B(u)$$

in the Skorohod space $D[0, 1]$, where $u \rightsquigarrow B(u)$ is a vector of independent standard Brownian motions, is immediate from Donsker's theorem. However, it does not seem so well known that the random process

$$u \rightsquigarrow \sqrt{n} \dot{\mathbb{M}}_n(u, \hat{\theta}_n) \quad \text{converges weakly to} \quad u \rightsquigarrow I(\theta_0)^{1/2} B^\circ(u) \quad (2)$$

in $D[0, 1]$, where $u \rightsquigarrow B^\circ(u)$ is a vector of independent standard Brownian bridges. Horváth and Parzen (1994) is apparently the first to introduce the statistic

$$\mathcal{T}_n = n \sup_{u \in [0, 1]} \dot{\mathbb{M}}_n(u, \hat{\theta}_n)^\top \hat{I}_n^{-1} \dot{\mathbb{M}}_n(u, \hat{\theta}_n)$$

for change point problems, where \hat{I}_n is a consistent estimator for the Fisher Information matrix $I(\theta_0)$. It is immediate from (2) and the continuous mapping theorem that

$$\mathcal{T}_n \rightarrow^d \sup_{u \in [0, 1]} \|B^\circ(u)\|^2.$$

Let us call this approach pioneered by Horváth and Parzen (1994) “ Z -process method”.

Although Horváth and Parzen (1994) didn't discuss the asymptotic behavior of the test under the alternative, Negri and Nishiyama (2011) who took the Z -process method for an ergodic diffusion process model based on the continuous observation proved also the consistency of the test under an alternative which has sufficient generality. Negri and Nishiyama's (2011) argument for alternatives can be applied also to the case of independent data. In Section 3 of this paper, we will present a generalized version of Z -process method which works also for some cases where the limit of the test statistic under the null hypothesis is a functional of a “mixture” of standard Brownian motions, where the mixture process, which is something like a “partial process of Fisher information”, is random and time dependent. We will also develop the argument of Negri and Nishiyama (2011) for the consistency under alternatives in a more general way. Some new examples will be given.

1.3 Notations and examples

In the rest of this section, we shall list up some examples to which our results can be applied. In what follows, the parameter space Θ is a bounded, open, convex subset of \mathbb{R}^d , where d is a fixed, positive integer. The word “vector” always means “ d -dimensional real column vector”, and the word “matrix” does “ $d \times d$ real matrix”. The Euclidean norm is denoted by $\|v\| := \sqrt{\sum_{i=1}^d |v^{(i)}|^2}$ for a vector v where $v^{(i)}$ denotes the i -th component of v , and by $\|A\| := \sqrt{\sum_{i,j=1}^d |A^{(i,j)}|^2}$ for a matrix A where $A^{(i,j)}$ denotes the (i,j) -component of A . Note that $\|Av\| \leq \|A\| \cdot \|v\|$ and $\|AB\| \leq \|A\| \cdot \|B\|$ for vector v and matrices A, B . The notations v^\top and A^\top denote the transpose. We use also the notation $A \circ B$ defined by $(A \circ B)^{(i,j)} := A^{(i,j)} B^{(i,j)}$ for two matrices A, B (the Hadamard product). We denote by I_d the identity matrix. The notations \rightarrow^p and \rightarrow^d mean the convergence in probability and the convergence in distribution, as $n \rightarrow \infty$, respectively.

Example A: Ergodic diffusion process. Let $I = (l, r)$, where $-\infty \leq l < r \leq \infty$, be given. Let us consider an I -valued diffusion process $t \rightsquigarrow X_t$ which is the unique strong solution to the stochastic differential equation (SDE)

$$X_t = X_0 + \int_0^t S(X_s; \alpha) ds + \int_0^t \sigma(X_s; \beta) dW_s,$$

where $s \rightsquigarrow W_s$ is a standard Wiener process. The parameters come from $\alpha \in \Theta_A \subset \mathbb{R}^{d_A}$ and $\beta \in \Theta_B \subset \mathbb{R}^{d_B}$, and we denote $\theta = (\alpha^\top, \beta^\top)^\top$. We are supposed to be able to observe the process X at discrete time grids $0 = t_0^n < t_1^n < \dots < t_n^n$, and we shall consider the asymptotic scheme $n\Delta_n^2 \rightarrow 0$ and $t_n^n \rightarrow \infty$ as $n \rightarrow \infty$, where

$$\Delta_n = \max_{1 \leq k \leq n} |t_k^n - t_{k-1}^n|,$$

and

$$\sum_{k=1}^n \left| \frac{|t_k^n - t_{k-1}^n|}{t_n^n} - \frac{1}{n} \right| \rightarrow 0, \quad \text{as } n \rightarrow \infty. \quad (3)$$

For Themes I and II, we introduce

$$\dot{\mathbb{M}}_n(\theta) = R_n^{-2} \dot{\mathbb{L}}_n(1, \theta) \quad \text{and} \quad \dot{\mathbb{M}}_n(u, \theta) = R_n^{-2} \dot{\mathbb{L}}_n(u, \theta),$$

respectively, where

$$\begin{aligned} & \mathbb{L}_n(u, \theta) \\ &= - \sum_{k: t_{k-1}^n \leq ut_n^n} \left\{ \log \sigma(X_{t_{k-1}^n}; \beta) + \frac{|X_{t_k^n} - X_{t_{k-1}^n} - S(X_{t_{k-1}^n}; \alpha)|^2 |t_k^n - t_{k-1}^n|}{2\sigma(X_{t_{k-1}^n}; \beta)^2 |t_k^n - t_{k-1}^n|} \right\} \end{aligned}$$

and R_n is the diagonal matrix such that $R_n^{(i,i)}$ is $\sqrt{t_n^n}$ for $i = 1, \dots, d_A$ and \sqrt{n} for $i = d_A + 1, \dots, d$ with $d = d_A + d_B$.

The problem to establish the moment convergence for M -estimators in this model, where X is a multi-dimensional diffusion process, was considered by Yoshida (2011), and Uchida and Yoshida (2012) relaxed the assumption $n\Delta_n^2 \rightarrow 0$ up to $n\Delta_n^a \rightarrow 0$, where $a \geq 2$ is a constant depending on the smoothness of the model as was done by Kessler (1995, 1997). To treat the parameters with different rates of convergence, in both papers an iterative method is used, and the method leads to some “adaptive estimators” that have advantages in applications as their simulation results show. In order to explain our core idea clearly, we only consider the one-dimensional diffusion process X under the sampling scheme $n\Delta_n^2 \rightarrow 0$. Some extension with no iterative argument to the case considered in Uchida and Yoshida (2012) could be possible.

Regarding Theme 2, Song and Lee (2009) proposed a statistic for testing the existence of a change point of the parameter β , but the problem to test it for both parameters was left as an open problem in their paper (see their Section 5). We will give an answer to this problem in Section 4.2.

Example B: Volatility of diffusion process. Let $I = (l, r)$, where $-\infty \leq l < r \leq \infty$, be given. Let us consider an I -valued diffusion process $t \rightsquigarrow X_t$ which is the unique strong solution to the SDE

$$X_t = X_0 + \int_0^t S(X_s)ds + \int_0^t \sigma(X_s; \theta)dW_s,$$

where $s \rightsquigarrow W_s$ is a standard Wiener process. Here, the drift coefficient $S(\cdot)$ is treated as an unknown nuisance function. We are supposed to be able to observe the process X at discrete time grids $0 = t_0^n < t_1^n < \dots < t_n^n = T < \infty$, and we shall consider the asymptotic scheme (3).

We introduce

$$\dot{\mathbb{M}}_n(\theta) = \frac{1}{n}\dot{\mathbb{L}}_n(1, \theta) \quad \text{and} \quad \dot{\mathbb{M}}_n(u, \theta) = \frac{1}{n}\dot{\mathbb{L}}_n(u, \theta)$$

for Themes I and II, respectively, where

$$\mathbb{L}_n(u, \theta) = - \sum_{k: t_{k-1}^n \leq ut_n^n} \left\{ \log \sigma(X_{t_{k-1}^n}; \theta) + \frac{|X_{t_k^n} - X_{t_{k-1}^n}|^2}{2\sigma(X_{t_{k-1}^n}; \theta)^2 |t_k^n - t_{k-1}^n|} \right\}.$$

The rate matrix is given by $R_n = \sqrt{n}I_d$.

Iacus and Yoshida (2012) proposed an estimator for the change point in a similar model. Our result for Theme II here, dealing with testing the existence of a change point, can be applied before statisticians proceed to their theory of estimation.

As we already mentioned, Song and Lee (2009) proposed a statistic for testing the existence of a change point in the volatility of an ergodic diffusion process under the asymptotic scheme $n\Delta_n^a \rightarrow 0$ and $n\Delta_n^b \rightarrow \infty$ for some $a > b > 4$, by an approach which is different from ours.

Example C: Cox's regression model. Let a sequence of counting processes $t \rightsquigarrow N_t^k$, $k = 1, 2, \dots$, which do not have simultaneous jumps, be observed during the time interval $[0, T]$. Suppose that $t \rightsquigarrow N_t^k$ has the intensity

$$\lambda_t^k(\theta) = \alpha(t)e^{\theta^\top Z_t^k} Y_t^k,$$

where the baseline hazard function α which is common for all k 's is non-negative and satisfies $\int_0^T \alpha(t)dt < \infty$, the random process $t \rightsquigarrow Z_t^k$ is a vector valued covariate for the individual k , and the random process $t \rightsquigarrow Y_t^k$ is given by

$$Y_t^k = \begin{cases} 1, & \text{if the individual } k \text{ is observed at time } t, \\ 0, & \text{otherwise.} \end{cases}$$

This model was introduced by Cox (1972), and its asymptotic theory was developed by Andersen and Gill (1982).

For Themes I and II, we introduce

$$\dot{\mathbb{M}}_n(\theta) = \frac{1}{n} \dot{\mathbb{L}}_n(1, \theta) \quad \text{and} \quad \dot{\mathbb{M}}_n(u, \theta) = \frac{1}{n} \dot{\mathbb{L}}_n(u, \theta),$$

respectively, where

$$\mathbb{L}_n(u, \theta) = \sum_{k=1}^n \int_0^{uT} (\theta^\top Z_t^k - \log S_t^{n,0}(\theta)) dN_t^k$$

with

$$S_t^{n,0}(\theta) = \sum_{k=1}^n e^{\theta^\top Z_t^k} Y_t^k.$$

The rate matrix is $R_n = \sqrt{n}I_d$.

Example D: Counting process models. Let $t \rightsquigarrow N_t$ be a counting process with the intensity $t \rightsquigarrow \lambda_t(\theta)$. We suppose that we can observe the processes on the compact time interval $[0, T_n]$, and consider the asymptotic scheme $T_n \rightarrow \infty$. Our results may be typically applied to

$$\dot{\mathbb{M}}_n(\theta) = \frac{1}{T_n} \dot{\mathbb{L}}_n(1, \theta) \quad \text{and} \quad \dot{\mathbb{M}}_n(u, \theta) = \frac{1}{T_n} \dot{\mathbb{L}}_n(u, \theta),$$

where

$$\mathbb{L}_n(u, \theta) = \int_0^{uT_n} \log \lambda_t(\theta) dN_t - \int_0^{uT_n} \lambda_t(\theta) dt.$$

The rate matrix is typically $R_n = \sqrt{T_n}I_d$. An example of this model is the stress release process introduced by Isham and Westcott (1979):

$$\lambda_t(\alpha, \beta, \theta) = \phi(\alpha t - \beta N_{t-}; \theta).$$

It is known that the process $t \rightsquigarrow X_t := \alpha t - \beta N_{t-}$ is ergodic under some mild conditions. A test statistic for the change point problem in this model, which is different from ours, has been proposed by Fujii and Nishiyama (2011).

Example E: Non-linear time series models. Let us consider the time series models of the form

$$X_k = a(X_{k-1}, X_{k-2}, \dots; \theta) + b(X_{k-1}, X_{k-2}, \dots; \theta)\varepsilon_k, \quad k = 1, 2, \dots$$

Here, $\{\varepsilon_k\}$ is an i.i.d. sequence with $E[\varepsilon_1] = 0$ or, more generally, a martingale difference sequence with respect to the filtration $(\mathcal{F}_k)_{k \geq 0}$ where $\mathcal{F}_k = \sigma(X_k, X_{k-1}, \dots)$. A possible way to define the estimating functions is

$$\dot{\mathbb{M}}_n(\theta) = \frac{1}{n} \dot{\mathbb{L}}_n(1, \theta) \quad \text{and} \quad \dot{\mathbb{M}}_n(u, \theta) = \frac{1}{n} \dot{\mathbb{L}}_n(u, \theta),$$

where

$$\mathbb{L}_n(u, \theta) = - \sum_{k=1}^{[un]} \left\{ \log b(X_{k-1}, X_{k-2}, \dots; \theta) + \frac{|X_k - a(X_{k-1}, X_{k-2}, \dots; \theta)|^2}{2b(X_{k-1}, X_{k-2}, \dots; \theta)^2} \right\}.$$

The rate matrix is typically given by $R_n = \sqrt{n}I_d$.

Some detailed discussions for Examples A, B, and C will be given in Sections 4, 5 and 6, respectively, while Examples D and E are not discussed in detail here.

2 Moment convergence of Z -estimators

Let Θ be a bounded, open, convex subset of \mathbb{R}^d . Let us be given a real valued random function $\mathbb{L}_n(\theta)$ of $\theta \in \Theta$ which is twice continuously differentiable with the gradient vector $\dot{\mathbb{L}}_n(\theta)$ and the Hessian matrix $\ddot{\mathbb{L}}_n(\theta)$, defined on a probability space (Ω, \mathcal{F}, P) that is common for all $n \in \mathbb{N}$. (However, it will be clear from our proofs that if the limit matrices $V(\theta_0)$ and $\ddot{M}(\theta)$ appearing below are non-random then the underlying probability spaces need not to be common for all $n \in \mathbb{N}$.) Let R_n be a (possibly, random) diagonal matrix whose diagonal components are positive, and define Q_n by $Q_n^{(i,j)} = (R_n^{(i,i)} R_n^{(j,j)})^{-1}$. Using these matrices, put

$$\dot{\mathbb{M}}_n(\theta) := R_n^{-2} \dot{\mathbb{L}}_n(\theta) \quad \text{and} \quad \ddot{\mathbb{M}}_n(\theta) := Q_n \circ \ddot{\mathbb{L}}_n(\theta). \quad (4)$$

(In the typical cases, $R_n = \sqrt{n}I_d$ and $Q_n = n^{-1}\mathbf{1}$, where $\mathbf{1}$ denotes the matrix whose all components are 1.)

First, we state a theorem to give an asymptotic representation for Z -estimators. Although this result is not really novel, we will give a proof for references.

Theorem 2.1 *Consider the above setting. Suppose that there exists a sequence of matrices $V_n(\theta_0)$ which are regular almost surely such that for any sequence of Θ -valued random vectors $\tilde{\theta}_n$ converging in probability to θ_0 ,*

$$\ddot{\mathbb{M}}_n(\tilde{\theta}_n) + V_n(\theta_0) \rightarrow^P 0.$$

Suppose also that

$$(R_n \dot{\mathbb{M}}_n(\theta_0), V_n(\theta_0)) \rightarrow^d (L(\theta_0), V(\theta_0)),$$

where $L(\theta_0)$ is a random vector and $V(\theta_0)$ is a random matrix which is regular almost surely (we do not assume that $V(\theta_0)$ and $L(\theta_0)$ are independent).

Then, for any sequence of Θ -valued random vectors $\widehat{\theta}_n$ which converges in probability to θ_0 and satisfies $\|R_n \dot{\mathbb{M}}_n(\widehat{\theta}_n)\| = o_P(1)$, it holds that

$$\begin{aligned} R_n(\widehat{\theta}_n - \theta_0) &= V_n(\theta_0)^{-1} R_n \dot{\mathbb{M}}_n(\theta_0) + o_P(1) \\ &\rightarrow^d V(\theta_0)^{-1} L(\theta_0). \end{aligned}$$

Remark. In this theorem, the consistency of the sequence of Z -estimators $\widehat{\theta}_n$ has been assumed. A method to show this property will be given in Lemma 3.1 (i) below.

Now, we give a theorem to establish the moment convergence of Z -estimators, which is the main result of this section.

Theorem 2.2 *Consider the setting described in the first paragraph of this section. Let some constants $p \geq 1$ and $a, b > 1$ such that $\frac{1}{a} + \frac{1}{b} = 1$ be given; see a remark at the end of the theorem for the case where we may set $a = 1$.*

Suppose that

$$\|R_n \dot{\mathbb{M}}_n(\theta_0)\| \text{ is asymptotically } L_{pa}\text{-bounded.} \quad (5)$$

Suppose also that there exist a constant $\gamma \in (0, 1]$ and some random matrices $\ddot{M}(\theta)$ indexed by $\theta \in \Theta$ such that

$$\lim_{n \rightarrow \infty} E \left[\sup_{\theta \in \Theta} \|R_n^\gamma(\ddot{\mathbb{M}}_n(\theta) - \ddot{M}(\theta))\|^{pa/\gamma} \right] = 0. \quad (6)$$

Suppose further that either of the following [M1] or [M2] is satisfied:

[M1] *There exists a random matrix $J(\theta_0)$ which is positive definite almost surely such that $\ddot{M}(\theta) \leq -J(\theta_0)$ for all $\theta \in \Theta$, almost surely, and that $E[\|J(\theta_0)^{-1}\|^{pb/\gamma}] < \infty$;*

[M2] *$E[\sup_{\theta \in \Theta} \|\ddot{M}(\theta)^{-1}\|^{pb/\gamma}] < \infty$, where the random matrices $\ddot{M}(\theta)$'s are assumed to be regular almost surely.*

Then, for any sequence of Θ -valued random vectors $\widehat{\theta}_n$ such that $\|R_n \dot{\mathbb{M}}_n(\widehat{\theta}_n)\|$ is asymptotically L_{pa} -bounded, it holds that $\|R_n(\widehat{\theta}_n - \theta_0)\|$ is asymptotically L_p -bounded. Therefore, in this situation, whenever we also have that $R_n(\widehat{\theta}_n - \theta_0) \rightarrow^d G(\theta_0)$ where $G(\theta_0)$ is a random vector, it holds for any continuous function ψ satisfying (1) for $q \in (0, p)$ that

$$\lim_{n \rightarrow \infty} E[\psi(R_n(\widehat{\theta}_n - \theta_0))] = E[\psi(G(\theta_0))],$$

where the limit is also finite.

When the last condition in [M1] is satisfied with $\|J(\theta_0)\|^{-1}$ which is bounded or the first condition in [M2] is satisfied with $\sup_{\theta \in \Theta} \|\ddot{M}(\theta)^{-1}\|$ which is bounded, the constant a appearing in the above claim may be replaced by 1.

Remark. The condition [M1] corresponds to the case $\rho = 2$ of the conditions [A3] and [A5] in Yoshida (2011), which are

$$M(\theta) - M(\theta_0) \leq -\chi(\theta_0)\|\theta - \theta_0\|^\rho, \quad \forall \theta \in \Theta,$$

and high order moment conditions on the positive random variable $\chi(\theta_0)^{-1}$, where “ $M(\theta)$ ” should be read as “ $Y(\theta)$ ” in Yoshida’s (2011) notation.

Proof of Theorem 2.1. Recalling (4), it follows from the Taylor expansion that

$$(R_n \dot{\mathbb{M}}_n(\widehat{\theta}_n))^{(i)} = (R_n \dot{\mathbb{M}}_n(\theta_0))^{(i)} + (\ddot{\mathbb{M}}_n(\widetilde{\theta}_n) R_n(\widehat{\theta}_n - \theta_0))^{(i)}, \quad i = 1, \dots, d, \quad (7)$$

where $\widetilde{\theta}_n$ is a random vector on the segment connecting θ_0 and $\widehat{\theta}_n$. So we have

$$R_n(\widehat{\theta}_n - \theta_0) = A_n + B_n R_n(\widehat{\theta}_n - \theta_0), \quad (8)$$

where

$$\begin{aligned} A_n &= V_n(\theta_0)^{-1} R_n(\dot{\mathbb{M}}_n(\theta_0) - \dot{\mathbb{M}}_n(\widehat{\theta}_n)), \\ B_n &= V_n(\theta_0)^{-1} (\ddot{\mathbb{M}}_n(\widetilde{\theta}_n) + V_n(\theta_0)). \end{aligned}$$

It follows from the extended continuous mapping theorem (e.g., Theorem 1.11.1 of van der Vaart and Wellner (1996)) that $V_n(\theta_0)^{-1} \xrightarrow{p} V(\theta_0)^{-1}$, thus we have $\|A_n\| = O_P(1)$ and $\|B_n\| = o_P(1)$. It therefore holds that

$$\|R_n(\widehat{\theta}_n - \theta_0)\| \leq O_P(1) + o_P(1) \cdot \|R_n(\widehat{\theta}_n - \theta_0)\|,$$

which implies that $\|R_n(\widehat{\theta}_n - \theta_0)\| = O_P(1)$. Hence, going back to (8) we obtain

$$R_n(\widehat{\theta}_n - \theta_0) = A_n + o_P(1) = V_n(\theta_0)^{-1} R_n \dot{\mathbb{M}}_n(\theta_0) + o_P(1).$$

The last claim is also a consequence of the extended continuous mapping theorem. The proof is finished. \square

Proof of Theorem 2.2. We will give a proof for the case where [M1] is assumed. The proof for the case where [M2] is assumed is similar (and simpler), so it is omitted.

Due to (7) again, we have

$$R_n(\widehat{\theta}_n - \theta_0) = C_n + (D_n^{(1)} + D_n^{(2)}) R_n(\widehat{\theta}_n - \theta_0),$$

where

$$\begin{aligned} C_n &= J(\theta_0)^{-1} R_n(\dot{\mathbb{M}}_n(\theta_0) - \dot{\mathbb{M}}_n(\widehat{\theta}_n)), \\ D_n^{(1)} &= J(\theta_0)^{-1} (\ddot{\mathbb{M}}_n(\widetilde{\theta}_n) - \ddot{M}(\widetilde{\theta}_n)), \\ D_n^{(2)} &= J(\theta_0)^{-1} (\ddot{M}(\widetilde{\theta}_n) + J(\theta_0)), \end{aligned}$$

where $\tilde{\theta}_n$ is a random vector on the segment connecting θ_0 and $\hat{\theta}_n$.

From now on, we consider the case $\gamma \in (0, 1)$; the proof for the case $\gamma = 1$ is easier, and it is omitted. Since $-D_n^{(2)}$ is non-negative definite almost surely, it follows from Minkowski's and Hölder's inequalities that

$$\begin{aligned} & (E[||R_n(\hat{\theta}_n - \theta_0)||^p])^{1/p} \\ & \leq (E[||(I_d - D_n^{(2)})R_n(\hat{\theta}_n - \theta_0)||^p])^{1/p} \\ & \leq (E[||C_n||^p])^{1/p} + (E[||R_n^\gamma D_n^{(1)}||^{p/\gamma}])^{\gamma/p} (E[||R_n^{1-\gamma}(\hat{\theta}_n - \theta_0)||^{p/(1-\gamma)}])^{(1-\gamma)/p} \\ & \leq O(1) + o(1) \cdot (E[||R_n^{1-\gamma}(\hat{\theta}_n - \theta_0)||^{p/(1-\gamma)}])^{(1-\gamma)/p}, \end{aligned}$$

where we have used Hölder's inequality again to get

$$E[||C_n||^p] \leq (E[||J(\theta_0)^{-1}||^{pb}])^{1/b} (E[||R_n(\dot{\mathbb{M}}_n(\theta_0) - \dot{\mathbb{M}}_n(\hat{\theta}_n))||^{pa})^{1/a}$$

and

$$E[||R_n^\gamma D_n^{(1)}||^{p/\gamma}] \leq (E[||J(\theta_0)^{-1}||^{pb/\gamma}])^{1/b} (E[||R^\gamma(\ddot{\mathbb{M}}_n(\tilde{\theta}_n) - \ddot{M}(\tilde{\theta}_n))||^{pa/\gamma})^{1/a};$$

if $||J(\theta_0)||^{-1}$ is bounded, we can get this kind of bounds with $a = 1$.

Notice that

$$\begin{aligned} & ||R_n^{1-\gamma}(\hat{\theta}_n - \theta_0)||^{1/(1-\gamma)} \\ & \leq \sqrt{d^{(1/(1-\gamma))-1} \sum_{i=1}^d |R_n^{(i,i)}|^2 |\hat{\theta}_n^{(i)} - \theta_0^{(i)}|^{2/(1-\gamma)}} \\ & \leq ||R_n(\hat{\theta}_n - \theta_0)|| \cdot d^{1/(2-2\gamma)} \cdot |\mathcal{D}(\Theta)|^{\gamma/(1-\gamma)}, \end{aligned}$$

where $\mathcal{D}(\Theta)$ denotes the diameter of Θ . So we obtain

$$\begin{aligned} & (E[||R_n(\hat{\theta}_n - \theta_0)||^p])^{1/p} \\ & \leq O(1) + o(1) \cdot (E[||R_n(\hat{\theta}_n - \theta_0)||^p])^{(1-\gamma)/p} \\ & \leq O(1) + o(1) \cdot (E[||R_n(\hat{\theta}_n - \theta_0)||^p] \vee 1)^{1/p}, \end{aligned}$$

which yields that

$$E[||R_n(\hat{\theta}_n - \theta_0)||^p] \leq O(1) + o(1) \cdot E[||R_n(\hat{\theta}_n - \theta_0)||^p].$$

Therefore, $||R_n(\hat{\theta}_n - \theta_0)||$ is asymptotically L_p -bounded. \square

3 Z-process method for change point problems

Let $D[0, 1]$ be the space of functions defined on $[0, 1]$ taking values in a finite-dimensional Euclidean space, which are right continuous and have left hand limits; we equip this space with the Skorohod metric. Throughout this section, all random processes, denoted as $u \rightsquigarrow X(u)$, are assumed to take values in $D[0, 1]$.

Let Θ be a bounded, open, convex subset of \mathbb{R}^d . For every $n \in \mathbb{N}$, let $u \rightsquigarrow \mathbb{L}_n(u, \theta)$ be a real valued random process indexed by $\theta \in \Theta$, defined on a probability space (Ω, \mathcal{F}, P) that is common for all $n \in \mathbb{N}$. (However, it will be clear from our proofs that the underlying probability spaces do not have to be common for $n \in \mathbb{N}$ if the objects $\dot{M}_{\theta_0}(u, \theta)$, $\dot{\mathcal{M}}(u, \theta)$ and $V(u, \theta_0)$ appearing in the limit below are non-random.) We suppose that for every $u \in [0, 1]$ the random function $\theta \mapsto \mathbb{L}_n(u, \theta)$ is two times continuously differentiable with the gradient vector $\dot{\mathbb{L}}_n(u, \theta)$ and the Hessian matrix $\ddot{\mathbb{L}}_n(u, \theta)$. Let a (possibly, random) diagonal matrix R_n whose diagonal components are positive be given, and define Q_n by $Q_n^{(i,j)} = (R_n^{(i,i)} R_n^{(j,j)})^{-1}$. Using these matrices, put

$$\dot{\mathbb{M}}_n(u, \theta) := R_n^{-2} \dot{\mathbb{L}}_n(u, \theta) \quad \text{and} \quad \ddot{\mathbb{M}}_n(u, \theta) := Q_n \circ \ddot{\mathbb{L}}_n(u, \theta).$$

We consider the following testing problem:

H_0 : the true value $\theta_0 \in \Theta$ does not change during $u \in [0, 1]$;

H_1 : “not H_0 ”.

The meaning of “not H_0 ” will be precisely specified in the condition [A] below. Let us describe some properties which the “limits” $\dot{M}_{\theta_0}(u, \theta)$ and $\dot{\mathcal{M}}(u, \theta)$ of the random vectors $\dot{\mathbb{M}}_n(u, \theta)$ under H_0 and under H_1 , respectively, have to satisfy.

[N] Under H_0 , it holds that

$$\sup_{\theta \in \Theta} \|\dot{\mathbb{M}}_n(1, \theta) - \dot{M}_{\theta_0}(1, \theta)\| \rightarrow^p 0, \quad (9)$$

where the limits $\dot{M}_{\theta_0}(1, \theta)$ ’s satisfy that

$$\inf_{\theta: \|\theta - \theta_0\| > \varepsilon} \|\dot{M}_{\theta_0}(1, \theta)\| > 0 = \|\dot{M}_{\theta_0}(1, \theta_0)\|, \quad \text{almost surely,} \quad \forall \varepsilon > 0. \quad (10)$$

[A] Under H_1 , it holds that

$$\sup_{u \in [0, 1]} \sup_{\theta \in \Theta} \|\dot{\mathbb{M}}_n(u, \theta) - \dot{\mathcal{M}}(u, \theta)\| \rightarrow^p 0, \quad (11)$$

where the limits $\dot{\mathcal{M}}(u, \theta)$ ’s satisfy that there exists a Θ -valued random vector θ_* such that

$$\inf_{\theta: \|\theta - \theta_*\| > \varepsilon} \|\dot{\mathcal{M}}(1, \theta)\| > 0 = \|\dot{\mathcal{M}}(1, \theta_*)\|, \quad \text{almost surely,} \quad \forall \varepsilon > 0, \quad (12)$$

and that

$$\sup_{u \in (0, 1)} \|\dot{\mathcal{M}}(u, \theta_*)\| > 0, \quad \text{almost surely.} \quad (13)$$

Assuming the conditions (9), (10), (11) and (12) is natural; see e.g. Theorems 5.7 and 5.9 of van der Vaart (1998). Let us explain how to check (13) in the most typical form of alternatives in the change problems:

H'_1 : there exists a constant $u_* \in (0, 1)$ such that the true value is $\theta_0 \in \Theta$ for $u \in [0, u_*]$, and $\theta_1 \in \Theta$ for $u \in (u_*, 1]$, where $\theta_0 \neq \theta_1$.

In many cases of “ergodic models”, under H'_1 , the condition (11) is satisfied with $\dot{\mathcal{M}}(u, \theta)$ such that

$$\dot{\mathcal{M}}(u_*, \theta) = u_* \dot{M}_{\theta_0}(1, \theta) \quad \text{and} \quad \dot{\mathcal{M}}(1, \theta) = u_* \dot{M}_{\theta_0}(1, \theta) + (1 - u_*) \dot{M}_{\theta_1}(1, \theta),$$

where not only $\dot{M}_{\theta_0}(1, \theta)$'s but also $\dot{M}_{\theta_1}(1, \theta)$'s are assumed to satisfy (10) with trivial change of notation. To see that the condition (13) is satisfied, notice that

$$\dot{\mathcal{M}}(u_*, \theta_*) = \dot{\mathcal{M}}(u_*, \theta_*) - u_* \dot{\mathcal{M}}(1, \theta_*) = u_*(1 - u_*)(\dot{M}_{\theta_0}(1, \theta_*) - \dot{M}_{\theta_1}(1, \theta_*));$$

if this were zero with positive probability, then it should follow from $\dot{\mathcal{M}}(1, \theta_*) = 0$ that $\dot{M}_{\theta_0}(1, \theta_*) = \dot{M}_{\theta_1}(1, \theta_*) = 0$ with positive probability, and this contradicts with (10) and the assumption that $\theta_0 \neq \theta_1$. Therefore, we have

$$\begin{aligned} \sup_{u \in (0,1)} \|\dot{\mathcal{M}}(u, \theta_*)\| &\geq \|\dot{\mathcal{M}}(u_*, \theta_*)\| \\ &= u_*(1 - u_*)\|\dot{M}_{\theta_0}(\theta_*) - \dot{M}_{\theta_1}(\theta_*)\| > 0, \quad \text{almost surely.} \end{aligned}$$

This positive value is closely related to the power of our test under H'_1 .

Now, we prepare a lemma to prove the consistency of a sequence of Z -estimators. This lemma can be proved exactly in the same way as Theorems 5.7 and 5.9 of van der Vaart (1998), so the proof is omitted.

Lemma 3.1 (i) Under $[N]$, for any sequence of Θ -valued random vectors $\hat{\theta}_n$ such that $\|\dot{\mathbb{M}}_n(1, \hat{\theta}_n)\| = o_P(1)$, it holds that $\hat{\theta}_n \rightarrow^p \theta_0$.

(ii) Under $[A]$, for any sequence of Θ -valued random vectors $\hat{\theta}_n$ such that $\|\dot{\mathbb{M}}_n(1, \hat{\theta}_n)\| = o_P(1)$, it holds that $\hat{\theta}_n \rightarrow^p \theta_*$.

We are ready to state our main result of this section.

Theorem 3.2 Consider the above situation. Let $\hat{\theta}_n$ be any sequence of Θ -valued random vectors such that $\|R_n \dot{\mathbb{M}}_n(1, \hat{\theta}_n)\| = o_P(1)$ under H_0 and H_1 . Let $u \rightsquigarrow \hat{V}_n(u)$ be any sequence of matrix valued random processes, which are regular except for $u = 0$ almost surely, and it should be a uniformly consistent sequence of estimators for the non-negative definite matrix valued random process $u \rightsquigarrow V(u, \theta_0)$ appearing below under H_0 . Introduce the test statistic

$$\mathcal{T}_n = \sup_{u \in (0,1]} (R_n \dot{\mathbb{M}}_n(u, \hat{\theta}_n))^\top (u \hat{V}_n(u)^{-1}) R_n \dot{\mathbb{M}}_n(u, \hat{\theta}_n).$$

(i) Under $[N]$, suppose that there exists a sequence of matrix valued random processes $u \rightsquigarrow V_n(u, \theta_0)$ such that $V_n(1, \theta_0)$'s are regular almost surely and that for any sequence of Θ -valued random vectors $\tilde{\theta}_n(u)$ indexed by $u \in [0, 1]$ satisfying $\sup_{u \in [0,1]} \|\tilde{\theta}_n(u) - \theta_0\| \rightarrow^p 0$,

$$\sup_{u \in [0,1]} \|\ddot{\mathbb{M}}_n(u, \tilde{\theta}_n(u)) + V_n(u, \theta_0)\| \rightarrow^p 0. \quad (14)$$

Suppose also that

$$(R_n \dot{\mathbb{M}}_n(u, \theta_0), V_n(u, \theta_0)) \rightarrow^d ((u^{-1}V(u, \theta_0))^{1/2}B(u), V(u, \theta_0)), \quad \text{in } D[0, 1], \quad (15)$$

where $u \rightsquigarrow V(u, \theta_0)$ is a non-negative definite matrix valued random process such that $V(1, \theta_0)$ is positive definite almost surely, and $u \rightsquigarrow B(u)$ is a vector of independent standard Brownian motions; the value of the first vector of the limit in (15) at $u = 0$ should be read as zero. (In general we do not assume that $u \rightsquigarrow V(u, \theta_0)$ and $u \rightsquigarrow B(u)$ are independent.)

If $\sup_{u \in [0, 1]} \|\widehat{V}_n(u) - V(u, \theta_0)\| \rightarrow^p 0$, then it holds that

$$\mathcal{T}_n \rightarrow^d \sup_{u \in [0, 1]} \|B(u) - u^{1/2}V(u, \theta_0)^{1/2}V(1, \theta_0)^{-1/2}B(1)\|^2. \quad (16)$$

Therefore the test is asymptotically distribution free if $V(u, \theta_0) = uV(1, \theta_0)$ for every $u \in [0, 1]$, because the limit in this case is reduced to $\sup_{u \in [0, 1]} \|B^\circ(u)\|^2$ where $u \rightsquigarrow B^\circ(u) = B(u) - uB(1)$ is a vector of independent standard Brownian bridges. In the general case, if $u \rightsquigarrow V(u, \theta_0)$ and $u \rightsquigarrow B(u)$ are independent, then the limit in (16) is approximated by

$$\sup_{u \in [0, 1]} \|B(u) - u^{1/2}\widehat{V}_n(u)^{1/2}\widehat{V}_n(1)^{-1/2}B(1)\|^2,$$

whose approximate distribution can be computed by some computer simulations for the standard Brownian motions $u \rightsquigarrow B(u)$.

(ii) Under $[A]$, it holds for any random point \check{u} in $(0, 1)$ that

$$\mathcal{T}_n \geq \lambda(\check{u}R_n^2\widehat{V}_n(\check{u})^{-1}) \left\{ \|\dot{\mathcal{M}}(\check{u}, \theta_*)\|^2 + o_P(1) \right\},$$

where $\lambda(A)$ denotes the smallest eigenvalue of the random matrix A . Hence, if there exists a random point \check{u} in $(0, 1)$ such that $\|\dot{\mathcal{M}}(\check{u}, \theta_*)\| > 0$ almost surely and that $\lambda(R_n^2\widehat{V}_n(\check{u})^{-1})$ tends to ∞ in probability, then the test is consistent.

Remark. In the typical cases of ergodic models, the matrix $V(u, \theta_0)$ is actually $uI(\theta_0)$ where $I(\theta_0)$ is the Fisher information matrix. Hence $V(u, \theta_0) = uV(1, \theta_0)$ holds, and the result is reduced to the standard case.

Proof. First let us prove (i). By Lemma 3.1 (i) we know that $\widehat{\theta}_n$ is a consistent estimator for θ_0 under H_0 . So it follows from Theorem 2.1 that

$$\begin{aligned} & R_n \dot{\mathbb{M}}_n(u, \widehat{\theta}_n) \\ &= R_n \dot{\mathbb{M}}_n(u, \theta_0) + \ddot{\mathbb{M}}_n(u, \widetilde{\theta}_n(u))R_n(\widehat{\theta}_n - \theta_0) \\ &= R_n \dot{\mathbb{M}}_n(u, \theta_0) - V_n(u, \theta_0)V_n(1, \theta_0)^{-1}R_n \dot{\mathbb{M}}_n(1, \theta_0) + e_n(u) \\ &\rightarrow^d (u^{-1}V(u, \theta_0))^{1/2}(B(u) - u^{1/2}V(u, \theta_0)^{1/2}V(1, \theta_0)^{-1/2}B(1)), \quad \text{in } D[0, 1], \end{aligned}$$

where $\widetilde{\theta}_n(u)$ is a random vector on the segment connecting θ_0 and $\widehat{\theta}_n$, and the reminder terms $e_n(u)$ appearing above satisfy that $\sup_{u \in [0, 1]} \|e_n(u)\| \rightarrow^p 0$. As a result the claim (i) follows from the continuous mapping theorem.

The inequality in (ii) is proved as follows:

$$\begin{aligned}
\mathcal{T}_n &\geq (R_n \dot{\mathbb{M}}_n(\check{u}, \hat{\theta}_n))^\top (\check{u} \hat{V}_n(\check{u})^{-1}) R_n \dot{\mathbb{M}}_n(\check{u}, \hat{\theta}_n) \\
&= \dot{\mathbb{M}}_n(\check{u}, \hat{\theta}_n)^\top (\check{u} R_n^2 \hat{V}_n(\check{u})^{-1}) \dot{\mathbb{M}}_n(\check{u}, \hat{\theta}_n) \\
&\geq \lambda(\check{u} R_n^2 \hat{V}_n(\check{u})^{-1}) \|\dot{\mathbb{M}}_n(\check{u}, \hat{\theta}_n)\|^2 \\
&= \lambda(\check{u} R_n^2 \hat{V}_n(\check{u})^{-1}) \left\{ \|\dot{\mathcal{M}}(\check{u}, \theta_*)\|^2 + o_P(1) \right\}.
\end{aligned}$$

The proof is finished. \square

4 Example A: Ergodic diffusion process

Recall the description of Example A in Section 1.3, where the first d_A -components α of the parameter $\theta = (\alpha^\top, \beta^\top)^\top$ is involved in the drift coefficient, and the latter d_B -components β is in the diffusion coefficient. Recalling also the definition of the rate matrix R_n there, let us consider the $(d_A + d_B)$ -dimensional random vectors $\dot{M}_n(u, \theta)$ and the $(d_A + d_B) \times (d_A + d_B)$ -random matrices $\ddot{M}_n(u, \theta)$ given as follows:

$$\begin{aligned}
\dot{M}_n(u, \theta) &= (\dot{M}_n^A(u, \theta)^\top, \dot{M}_n^B(u, \theta)^\top)^\top, \\
\ddot{M}_n(u, \theta) &= \begin{pmatrix} \ddot{M}_n^A(u, \theta) & \ddot{M}_n^C(u, \theta) \\ \ddot{M}_n^C(u, \theta)^\top & \ddot{M}_n^B(u, \theta) \end{pmatrix}.
\end{aligned}$$

Below, we will use the following notation: for a given constant $p \geq 1$ and a given sequence of positive constants r_n ,

$$\xi_n = o_{M(p)}(r_n^{-1}) \iff r_n E[\|\xi_n\|^p] \rightarrow 0. \quad (17)$$

Notice that $\xi_n = o_{M(1)}(r_n^{-1})$ implies that $\xi_n = o_P(r_n^{-1})$.

Under some regularity conditions which are usually assumed in the asymptotic theory for ergodic diffusion process models, it is standard to show the following facts (see e.g. the appendix of Kessler (1997) for some techniques needed to prove them; see Nishiyama (2011), in Japanese, for the detailed proofs of the techniques that are omitted in Kessler's (1997) appendix):

$$\begin{aligned}
\sup_{u \in [0,1]} \left\| \dot{\mathbb{M}}_n^A(u, \theta_0) - \frac{1}{t_n} \sum_{k: t_{k-1}^n \leq ut_n} \frac{\dot{S}(X_{t_{k-1}^n}; \alpha_0)}{\sigma(X_{t_{k-1}^n}; \beta_0)} (W_{t_k^n} - W_{t_{k-1}^n}) \right\| &= o_{M(p)}((t_n^n)^{-1/2}), \\
\sup_{u \in [0,1]} \left\| \dot{\mathbb{M}}_n^B(u, \theta_0) - \frac{1}{n} \sum_{k: t_{k-1}^n \leq ut_n} \frac{\dot{\sigma}(X_{t_{k-1}^n}; \beta_0)}{\sigma(X_{t_{k-1}^n}; \beta_0)} \left\{ \frac{|W_{t_k^n} - W_{t_{k-1}^n}|^2}{|t_k^n - t_{k-1}^n|} - 1 \right\} \right\| &= o_{M(p)}(n^{-1/2}), \\
\sup_{u \in [0,1]} \sup_{\theta \in \Theta} \left\| \ddot{\mathbb{M}}_n^A(u, \theta) - \frac{1}{t_n} \sum_{k: t_{k-1}^n \leq ut_n} H^A(X_{t_{k-1}^n}; \theta_0, \theta) |t_k^n - t_{k-1}^n| \right\| &= o_{M(p)}((t_n^n)^{-1/2}),
\end{aligned}$$

$$\sup_{u \in [0,1]} \sup_{\theta \in \Theta} \left\| \ddot{\mathbb{M}}_n^B(u, \theta) - \frac{1}{n} \sum_{k: t_{k-1}^n \leq ut_n} H^B(X_{t_{k-1}^n}; \theta_0, \theta) \right\| = o_{M(p)}(n^{-1/2}),$$

$$\sup_{u \in [0,1]} \sup_{\theta \in \Theta} \|\ddot{\mathbb{M}}_n^C(u, \theta)\| = o_{M(p)}(n^{-1/4}),$$

where

$$H^A(x; \theta_0, \theta) = \frac{\ddot{S}(x; \alpha)(S(x; \alpha_0) - S(x; \alpha)) - \dot{S}(x; \alpha)\dot{S}(x; \alpha)^\top}{\sigma(x; \beta)^2},$$

$$H^B(x; \theta_0, \theta) = \left\{ \frac{\ddot{\sigma}(x; \beta)}{\sigma(x; \beta)^3} - 3 \frac{\dot{\sigma}(x; \beta)\dot{\sigma}(x; \beta)^\top}{\sigma(x; \beta)^4} \right\} (\sigma(x; \beta_0)^2 - \sigma(x; \beta)^2) - 2 \frac{\dot{\sigma}(x; \beta)\dot{\sigma}(x; \beta)^\top}{\sigma(x; \beta)^2}.$$

The regularity conditions for the above claims depend on the constant $p \geq 1$ appearing in “ $o_{M(p)}(n^{-1/2})$ ” which we need to have.

4.1 Moment convergence

The assumption (5) can be checked by applying Burkholder-Davis-Gundy’s inequality to the main part of $R_n \ddot{\mathbb{M}}_n(\theta_0) = R_n \ddot{\mathbb{M}}_n(1, \theta_0)$. On the other hand, noting also $t_n^n \leq n$, we can apply Remark 1 (ii) of Uchida and Yoshida (2012) to show that the assumption (6) for $\ddot{\mathbb{M}}_n(\theta) = \ddot{\mathbb{M}}_n(1, \theta)$ is satisfied for

$$\ddot{M}(\theta) = \begin{pmatrix} \ddot{M}^A(\theta) & 0 \\ 0 & \ddot{M}^B(\theta) \end{pmatrix},$$

with

$$\ddot{M}^A(\theta) = \int_I H^A(x; \theta_0, \theta) \mu_{\theta_0}(dx) \quad \text{and} \quad \ddot{M}^B(\theta) = \int_I H^B(x; \theta_0, \theta) \mu_{\theta_0}(dx),$$

where μ_{θ_0} denotes the invariant distribution of X when the true value is θ_0 . In order to make the assumption [M1] or [M2] fulfilled, we have to introduce the parametric model for the drift and diffusion coefficients nicely. An example for which the assumption [M1] can be easily checked is $S(\cdot; \alpha) = \alpha^\top a(\cdot)$ and $\sigma(\cdot; \beta) = e^{\beta^\top b(\cdot)}$, where $a(\cdot)$ and $b(\cdot)$ are some vectors of known functions, assuming that $b(\cdot)$ is bounded. The assumption [M2] would be satisfied in more general parametric models, because $\ddot{M}(\theta)$ ’s are non-random in this example.

4.2 Change point problem

Under some standard conditions on the parametric family for the drift and diffusion coefficients in the context of ergodic diffusion processes, we can show that

the condition (9) under H_0 is satisfied with $\dot{M}_{\theta_0}(1, \theta) = (\dot{M}_{\theta_0}^A(1, \theta)^\top, \dot{M}_{\theta_0}^B(1, \theta)^\top)^\top$, where

$$\dot{M}_{\theta_0}^A(1, \theta) = \int_I \frac{\dot{S}(x; \alpha)}{\sigma(x; \beta)} (S(x; \alpha_0) - S(x; \alpha)) \mu_{\theta_0}(dx)$$

and

$$\dot{M}_{\theta_0}^B(1, \theta) = \int_I \frac{\dot{\sigma}(x; \beta)}{\sigma(x; \beta)^3} (\sigma(x; \beta_0)^2 - \sigma(x; \beta)^2) \mu_{\theta_0}(dx),$$

and that the condition (11) under H'_1 is satisfied with

$$\dot{\mathcal{M}}(u, \theta) = (u \wedge u_*) \dot{M}_{\theta_0}(1, \theta) + ((u - u_*) \vee 0) \dot{M}_{\theta_1}(1, \theta).$$

As stated there, the condition (13) is automatically satisfied as soon as the natural conditions (10) and (12) are satisfied.

Using the facts which we presented at the beginning of this section and the usual martingale central limit theorem, we can see that the condition (14) and (15) hold for

$$V_n(u, \theta_0) = \begin{pmatrix} V_n^A(u, \theta_0) & 0 \\ 0 & V_n^B(u, \theta_0) \end{pmatrix},$$

where

$$\begin{aligned} V_n^A(u, \theta_0) &= \frac{1}{t_n^n} \sum_{k: t_{k-1}^n \leq ut_n^n} \frac{\dot{S}(X_{t_{k-1}^n}; \alpha_0) \dot{S}(X_{t_{k-1}^n}; \alpha_0)^\top}{\sigma(X_{t_{k-1}^n}; \beta_0)^2} |t_k^n - t_{k-1}^n|, \\ V_n^B(u, \theta_0) &= \frac{2}{n} \sum_{k: t_{k-1}^n \leq ut_n^n} \frac{\dot{\sigma}(X_{t_{k-1}^n}; \beta_0) \dot{\sigma}(X_{t_{k-1}^n}; \beta_0)^\top}{\sigma(X_{t_{k-1}^n}; \beta_0)^2}. \end{aligned}$$

The limit of $V_n(u, \theta_0)$ is $V(u, \theta_0) = u I_{\theta_0}(\theta_0)$, where

$$I_{\theta_0}(\theta) = \begin{pmatrix} I_{\theta_0}^A(\theta) & 0 \\ 0 & I_{\theta_0}^B(\theta) \end{pmatrix}$$

with

$$I_{\theta_0}^A(\theta) = \int_I \frac{\dot{S}(x; \alpha) \dot{S}(x; \alpha)^\top}{\sigma(x; \beta)^2} \mu_{\theta_0}(dx), \quad I_{\theta_0}^B(\theta) = 2 \int_I \frac{\dot{\sigma}(x; \beta) \dot{\sigma}(x; \beta)^\top}{\sigma(x; \beta)^2} \mu_{\theta_0}(dx).$$

We suppose that $I_{\theta_0}(\theta)$'s are positive definite.

As a consistent estimator $\hat{V}_n(u)$ for $V(u, \theta_0)$, we introduce

$$\hat{V}_n(u) = \begin{pmatrix} u \hat{I}_n^A & 0 \\ 0 & u \hat{I}_n^B \end{pmatrix},$$

where

$$\hat{I}_n^A = \frac{1}{n} \sum_{k=1}^n \frac{\dot{S}(X_{t_{k-1}^n}; \hat{\alpha}_n) \dot{S}(X_{t_{k-1}^n}; \hat{\alpha}_n)^\top}{\sigma(X_{t_{k-1}^n}; \hat{\beta}_n)^2}, \quad \hat{I}_n^B = \frac{2}{n} \sum_{k=1}^n \frac{\dot{\sigma}(X_{t_{k-1}^n}; \hat{\beta}_n) \dot{\sigma}(X_{t_{k-1}^n}; \hat{\beta}_n)^\top}{\sigma(X_{t_{k-1}^n}; \hat{\beta}_n)^2}.$$

Since $V(u, \theta_0) = uV(1, \theta_0)$ in this example, the limit of our test statistic \mathcal{T}_n is $\sup_{u \in [0,1]} \|B^\circ(u)\|^2$ where $u \rightsquigarrow B^\circ(u)$ is a vector of standard Brownian bridges, so the test is asymptotically distribution free.

Finally, it is clear that under H'_1 , $\lambda(R_n^2 \widehat{V}_n(u_*)^{-1})$ tends to ∞ in probability since the matrix $u_* I_{\theta_0}(\theta_*) + (1 - u_*) I_{\theta_1}(\theta_*)$, which is the limit of $\widehat{V}_n(u_*)$, is positive definite. Thus the test is consistent.

4.3 Numerical study for change point problem

In this section, as well as Section 5.3 for Example B, we observe finite sample performance of our test statistic through numerical experiments. Here, we adopt the Ornstein-Uhlenbeck process starting from $x_0 = 0$ for the true (data-generating) process:

$$X_t = x_0 - \int_0^t \alpha X_s ds + \beta W_t, \quad t \in [0, T]. \quad (18)$$

For simplicity, we shall treat the equidistant sampling case, that is, $\Delta_n = |t_k^n - t_{k-1}^n|$ for every $k = 1, \dots, n$.

We are going to observe the trajectory of the process (18) for different time horizons $t_n^n = T$, and the number n of observations for each trajectory is such that $t_n^n = n^{1/3}$, so $\Delta_n = n^{-2/3}$. For this process (18) the estimators for the parameters α and β and the estimator of the information matrix can be explicitly calculated, and thus the test statistic can be easily computed. For any fixed level $\varepsilon > 0$ the critical value c_ε is given by

$$P \left(\sup_{u \in [0,1]} \sum_{i=1}^{d_A+d_B} |B^{\circ,(i)}(u)|^2 > c_\varepsilon \right) = \varepsilon.$$

Table 1 of Lee *et al.* (2003) gives a table of the critical values for the significance levels $\varepsilon = 0.01, 0.05, 0.10$ and for different values of the dimension $d = d_A + d_B$ computed by Monte Carlo simulation for the limit distribution. Throughout we take the significance level to be $\varepsilon = 0.05$. For two parameters ($d = 2$) the critical value is $c_\varepsilon = 2.408$. Regarding the null hypothesis we generate $M = 10^4$ trajectory of (18) and we evaluate the empirical size. The results are reported in Table 1. We observe that: the empirical size gains along with increasing terminal time $T = t_n^n$, attaining at 0.05, but also for small terminal T . In the second example reported in Table 1, the values of the parameter are the maximum likelihood estimate for the mostly federal funds data 1963-1998 in Aït-Sahalia (1999).

Regarding the alternative hypothesis we study the behavior of the test statistic in three different situations and for different change point $u_* T$ of the parameters, as follows:

- The drift coefficient changes from α_0 to α_1 , but the diffusion coefficient does not change.
- The drift coefficient does not change, but the diffusion coefficient changes from β_0 to β_1 .

T	5	10	15	20	25
n	125	1000	3375	8000	15625
$\alpha_0 = 1, \beta_0 = 1$	0.044	0.054	0.050	0.052	0.053
$\alpha_0 = 0.25, \beta_0 = 0.02$	0.047	0.061	0.058	0.064	0.054

Table 1: Empirical size based on $M = 10^4$ independent statistics, for different time horizons.

- Both coefficients change.

For each of the above scenarios we consider the following change points, $u_* = \frac{1}{2}, \frac{3}{4}, \frac{9}{10}$.

The first scenario is the worst case for the diffusion, and in order to detect a change in the drift we have to observe the process as long as possible. Table 2 shows empirical power for different terminal times T and different change points u_*T . The values of the parameters are $\alpha_0 = 0.25$, $\alpha_1 = 0.50$ and $\beta = 0.02$. The last does not vary. We simulate 10^4 independent copies of a trajectory of (18) to obtain different values of \mathcal{T}_n . The power increase as T increase and the performance is better when we can observe the process after the change for long time (the case $u_* = \frac{1}{2}$). In such a case the power of the test is reasonable. In the worst case $u_* = \frac{9}{10}$, the test is not able to detect the change in the drift coefficient.

T	5	10	15	20	25
n	125	1000	3375	8000	15625
$u_* = \frac{1}{2}$	0.31	0.52	0.73	0.79	0.88
$u_* = \frac{3}{4}$	0.12	0.17	0.23	0.26	0.35
$u_* = \frac{9}{10}$	0.05	0.07	0.08	0.08	0.09

Table 2: Empirical power based on $M = 10^4$ independent statistics. Here the significance level is 0.05. The values of the parameter are $\alpha_0 = 0.25$, $\alpha_1 = 0.50$ and $\beta = 0.02$ (it does not vary).

Table 3 reports the results for simulation when only the drift changes, but the change is bigger. With $\alpha_0 = 0.25$, $\alpha_1 = 1.25$ and $\beta = 0.02$, the power increases

not only for $u_* = \frac{1}{2}$ but also for $u_* = \frac{3}{4}$. This was expected, but as in the previous example the performance of the test is not good when the change of the parameter occur at the end of the observation window.

T	5	10	15	20	25
n	125	1000	3375	8000	15625
$u_* = \frac{1}{2}$	0.35	0.60	0.78	0.88	0.94
$u_* = \frac{3}{4}$	0.13	0.20	0.28	0.31	0.38
$u_* = \frac{9}{10}$	0.06	0.08	0.09	0.11	0.11

Table 3: Empirical power based on $M = 10^4$ independent statistics. Here the significance level is 0.05. The values of the parameter are $\alpha_0 = 0.25$, $\alpha_1 = 1.25$ and $\beta = 0.02$ (it does not vary).

Table 4 shows the empirical power for different terminal times T and different change points u_* in the second scenario: $\alpha = 0.25$ does not vary, but $\beta_0 = 0.02$ changes and becomes $\beta_1 = 0.03$. The power of the test is very good, but this is not surprising because a change in the diffusion coefficient can be easily detected. The situation reported in Table 5 is the same: also for very small change in the diffusion coefficient, the performance of the test is very good with the empirical power that reaches the value 1 also for small T . We do not report the results for the third scenario where the drift changes at the same instant of the diffusion, because the performance of the test is the same as in the second scenario. This is not surprising and is due to the different rates of convergence of the estimators of the two parameters.

5 Example B: Volatility of diffusion process

Recall the description of Example B in Section 1.3. An interesting point of this example is that the limit of $-\dot{\mathbb{M}}_n(u, \tilde{\theta}_n(u))$ is random and depend on $u \in [0, 1]$ in a complex way.

Let a constant $p \geq 1$ be given, and recall the notation (17). Under some regularity conditions, it holds that

$$\sup_{u \in [0, 1]} \left\| \dot{\mathbb{M}}_n(u, \theta_0) - \frac{1}{n} \sum_{k: t_{k-1}^n \leq ut_n} \frac{\dot{\sigma}(X_{t_{k-1}^n}; \theta_0)}{\sigma(X_{t_{k-1}^n}; \theta_0)} \left\{ \frac{|W_{t_k^n} - W_{t_{k-1}^n}|^2}{|t_k^n - t_{k-1}^n|} - 1 \right\} \right\| = o_{M(p)}(n^{-1/2}),$$

T	5	10	15	20	25
n	125	1000	3375	8000	15625
$u_* = \frac{1}{2}$	0.99	1	1	1	1
$u_* = \frac{3}{4}$	0.86	1	1	1	1
$u_* = \frac{9}{10}$	0.36	0.99	1	1	1

Table 4: Empirical power based on $M = 10^4$ independent statistics. Here the significance level is 0.05. The values of the parameter are $\alpha = 0.25$ (it does not vary), $\beta_0 = 0.02$ and $\beta_1 = 0.03$.

$$\sup_{u \in [0,1]} \sup_{\theta \in \Theta} \left\| \ddot{\mathbb{M}}_n(u, \theta) - \frac{1}{n} \sum_{k: t_{k-1}^n \leq ut_n^n} H(X_{t_{k-1}^n}; \theta_0, \theta) \right\| = o_{M(p)}(n^{-1/2}),$$

where

$$\begin{aligned} H(x; \theta_0, \theta) = & \left\{ \frac{\ddot{\sigma}(x; \theta)}{\sigma(x; \theta)^3} - 3 \frac{\dot{\sigma}(x; \theta) \dot{\sigma}(x; \theta)^\top}{\sigma(x; \theta)^4} \right\} (\sigma(x; \theta_0)^2 - \sigma(x; \theta)^2) \\ & - 2 \frac{\dot{\sigma}(x; \theta) \dot{\sigma}(x; \theta)^\top}{\sigma(x; \theta)^2}. \end{aligned}$$

The regularity conditions for the above claims depend on the constant $p \geq 1$ which we need to have. Moreover, under some standard conditions, it holds that for any sequence of random vectors $\tilde{\theta}_n(u)$ indexed by $u \in [0, 1]$ such that $\sup_{u \in [0,1]} \|\tilde{\theta}_n(u) - \theta_0\| \rightarrow^p 0$,

$$\sup_{u \in [0,1]} \|\ddot{\mathbb{M}}_n(u, \tilde{\theta}_n(u)) + V_n(u, \theta_0)\| \rightarrow^p 0,$$

where

$$V_n(u, \theta_0) = \frac{2}{n} \sum_{k: t_{k-1}^n \leq ut_n^n} \frac{\dot{\sigma}(X_{t_{k-1}^n}; \theta_0) \dot{\sigma}(X_{t_{k-1}^n}; \theta_0)^\top}{\sigma(X_{t_{k-1}^n}; \theta_0)^2}, \quad \forall u \in [0, 1].$$

Also, it follows from the well known theory of martingales that

$$(\sqrt{n} \dot{\mathbb{M}}_n(u, \theta_0), V_n(u, \theta_0)) \rightarrow^d ((u^{-1} V(u, \theta_0))^{1/2} B(u), V(u, \theta_0)) \quad \text{in } D[0, 1],$$

where $u \rightsquigarrow B(u)$ is a vector of independent standard Brownian motions which is independent of the matrix valued random process $u \rightsquigarrow V(u, \theta_0)$ given by

$$V(u, \theta_0) = 2 \int_0^{uT} \frac{\dot{\sigma}(X_s; \theta_0) \dot{\sigma}(X_s; \theta_0)^\top}{\sigma(X_s; \theta_0)^2} ds, \quad \forall u \in [0, 1].$$

T	5	10	15	20	25
n	125	1000	3375	8000	15625
$u_* = \frac{1}{2}$	0.87	1	1	1	1
$u_* = \frac{3}{4}$	0.52	0.99	1	1	1
$u_* = \frac{9}{10}$	0.14	0.62	0.99	1	1

Table 5: Empirical power based on $M = 10^4$ independent statistics. Here the significance level is 0.05. The values of the parameter are $\alpha = 0.25$ (it does not vary), $\beta_0 = 0.020$ and $\beta_1 = 0.025$.

5.1 Moment convergence

Due to the above facts (for $u = 1$), Theorem 2.1 yields that for any consistent estimator $\hat{\theta}_n$ for θ_0 satisfying $\|\dot{\mathbb{M}}_n(\hat{\theta}_n)\| = o_P(n^{-1/2})$ we have $\sqrt{n}(\hat{\theta}_n - \theta_0) \rightarrow^d V(\theta_0)^{-1/2}Z$, where Z is a standard Gaussian random vector which is independent of $V(\theta_0) = V(1, \theta_0)$.

Next let us apply Theorem 2.2. The assumption (5) for $\sqrt{n}\dot{\mathbb{M}}_n(\theta_0) = \sqrt{n}\dot{\mathbb{M}}_n(1, \theta_0)$ can be checked by using Burkholder-Davis-Gundy's inequality. In the case of this example, checking that the assumption (6) for $\ddot{\mathbb{M}}_n(\theta) = \ddot{\mathbb{M}}_n(1, \theta)$ is satisfied with

$$\ddot{M}(\theta) = \int_0^T H(X_t; \theta_0, \theta) dt$$

is easy. In order to make the assumption [M1] or [M2] fulfilled, we again have to introduce the parametric model for the diffusion coefficients nicely. An example for which the former assumption in [M1] can be easily checked is $\sigma(\cdot; \theta) = e^{\theta^\top g(\cdot)}$, where $g(\cdot)$ are some vectors of known, bounded functions. The latter assumption in [M1] is then reduced to

$$E \left[\left\| \left(\int_0^T g(X_t) g(X_t)^\top dt \right)^{-1} \right\|^{pb/\gamma} \right] < \infty,$$

for which we can give a clear sufficient condition for the function g at least in the one-dimensional case (for example, just assume $|g(\cdot)|^2 \geq c$ for a constant $c > 0$).

5.2 Change point problem

Under some standard conditions on the parametric family for the diffusion coefficient, we can show that the condition (9) under H_0 is satisfied with

$$\dot{M}_{\theta_0}(1, \theta) = \int_0^{uT} \frac{\dot{\sigma}(X_t; \theta)}{\sigma(X_t; \theta)^3} (\sigma(X_t; \theta_0)^2 - \sigma(X_t; \theta)^2) dt.$$

Under H'_1 , we have

$$\sup_{u \in [0, 1]} \|\dot{M}_n(u, \theta) - \dot{M}(u, \theta)\| \rightarrow^p 0,$$

where

$$\begin{aligned} \dot{M}(u, \theta) &= \int_0^{(u \wedge u_*)T} \frac{\dot{\sigma}(X_t; \theta)}{\sigma(X_t; \theta)^3} (\sigma(X_t; \theta_0)^2 - \sigma(X_t; \theta)^2) dt \\ &\quad + 1\{u > u_*\} \int_{u_*T}^{uT} \frac{\dot{\sigma}(X_t; \theta)}{\sigma(X_t; \theta)^3} (\sigma(X_t; \theta_1)^2 - \sigma(X_t; \theta)^2) dt. \end{aligned}$$

We can give a set of sufficient conditions for (13) as follows. Suppose that the Lebesgue measure of the random set $T_0 = \{t \in [0, T] : \dot{\sigma}(X_t; \theta_*)/\sigma(X_t; \theta_*)^3 = 0\}$ is zero almost surely, which is true in many concrete models. In this case, replace the values $\dot{\sigma}(X_t; \theta_*)/\sigma(X_t; \theta_*)^3$ in the definition of $\dot{M}(u, \theta_*)$ on the set T_0 by 1 to construct $\dot{M}(u, \theta_*)^*$ which equals with the original $\dot{M}(u, \theta_*)$ for all $u \in [0, 1]$, almost surely. If we further assume that for any non-empty interval $J \subset I$

$$\sigma(x; \theta) = \sigma(x; \theta'), \quad \forall x \in J \iff \theta = \theta'$$

and that each of the random sets $J_0 = \{X_t : t \in [0, u_*T]\}$ and $J_1 = \{X_t : t \in (u_*T, T]\}$ includes a non-empty set almost surely, then it follows from the assumption $\theta_0 \neq \theta_1$ that $\frac{d}{du} \dot{M}(u, \theta_*)^* \not\equiv 0$ almost surely. Thus we have

$$\sup_{u \in (0, 1)} \|\dot{M}(u, \theta_*)\| = \sup_{u \in (0, 1)} \|\dot{M}(u, \theta_*)^*\| > 0, \quad \text{almost surely.}$$

Now, consider the matrices

$$V(u, \theta) = 2 \int_0^{uT} \frac{\dot{\sigma}(X_t; \theta) \dot{\sigma}(X_t; \theta)^\top}{\sigma(X_t; \theta)^2} dt, \quad \forall u \in [0, 1].$$

Let us assume that for every $\theta \in \Theta$ there exists a set $J_\theta \subset I$ such that the Lebesgue measure of J_θ^c is zero and that the matrices $\dot{\sigma}(x; \theta) \dot{\sigma}(x; \theta)^\top / \sigma(x; \theta)$ are positive definite for $x \in J_\theta$, which is a standard assumption. In this case, if the claim that the Lebesgue measure of the set $\{t \in [0, T]; X_t(\omega) \in J\}$ is positive for any set $J \subset I$ such that the Lebesgue measure of J^c is zero holds for almost all ω , then $V(u, \theta_0)$'s and $V(u, \theta_*)$'s for $u \in (0, 1]$ are positive definite almost surely under H_0 and H'_1 , respectively.

As we saw at the beginning of this section, the conditions (14) and (15) under H_0 are satisfied. As a consistent estimator $\widehat{V}_n(u)$ for $V(u, \theta_0)$ is given by

$$\widehat{V}_n(u) = \frac{2}{n} \sum_{k: t_{k-1}^n \leq ut_n^n} \frac{\dot{\sigma}(X_{t_{k-1}^n}^n; \widehat{\theta}_n) \dot{\sigma}(X_{t_{k-1}^n}^n; \widehat{\theta}_n)^\top}{\sigma(X_{t_{k-1}^n}^n; \widehat{\theta}_n)^2}, \quad \forall u \in [0, 1].$$

Our test in this example is not asymptotically distribution free.

Finally, it is clear that under H'_1 , $n\lambda(\widehat{V}_n(\tilde{u})^{-1})$ tends to ∞ in probability, because it follows from what we have assumed that $\lambda(\widehat{V}_n(\tilde{u})^{-1}) \rightarrow^p \lambda(V(\tilde{u}, \theta_*)^{-1})$ and the limit is positive almost surely. Thus the test is consistent.

5.3 Numerical study for change point problem

The data-generating process is the following:

$$X_t = 4 - \int_0^t (X_s - 4) ds + \int_0^t \exp\left(\theta \frac{X_s^2}{1 + X_s^2}\right) dW_s, \quad t \in [0, 1],$$

where the drift coefficient $S(x) = -(x - 4)$ is treated as a nuisance function. Suppose that we observe $M = 10^3$ independent copies of this process at the equidistant time grid $t_k^n = \frac{k}{n}$, $k = 0, 1, \dots, n$. We compute the critical value of the test based on the approximation of the limit distribution

$$\sup_{u \in [0, 1]} |B(u) - u^{1/2} V(u, \theta_0)^{1/2} V(1, \theta_0)^{-1/2} B(1)|^2 \quad (19)$$

obtained by replacing

$$V(u, \theta_0) = 2 \int_0^u \left| \frac{X_s^2}{1 + X_s^2} \right|^2 ds$$

by the natural estimator

$$\widehat{V}_n(u) = \frac{2}{n} \sum_{k=1}^{[un]} \left| \frac{X_{t_{k-1}^n}^2}{1 + X_{t_{k-1}^n}^2} \right|^2$$

and doing 10^3 times Monte Carlo simulation for the standard Brownian motion $u \rightsquigarrow B(u)$.

The empirical size under H_0 is reported in Table 6, where the true value of the parameter is set as $\theta_0 = 1.0$ or 1.5 . We see that the convergence to the approximate distribution of (19) is not perfectly good, but it is reasonable even for the cases where n is small.

The empirical power under H'_1 is reported in Table 7, where the true values of the parameter change from $\theta_0 = 1.0$ to $\theta_1 = 1.5$ at time point $u_* = \frac{1}{2}, \frac{3}{4}$ or $\frac{9}{10}$.

n	20	40	100	200
$\theta_0 = 1.0$	0.026	0.024	0.042	0.040
$\theta_0 = 1.5$	0.026	0.023	0.037	0.034

Table 6: Empirical size based on $M = 10^3$ independent statistics. Here the significance level is 0.05. The value of the parameter is $\theta_0 = 1.0$ or 1.5.

n	20	40	100	200
$u_* = \frac{1}{2}$	0.067	0.331	0.755	0.946
$u_* = \frac{3}{4}$	0.125	0.255	0.630	0.873
$u_* = \frac{9}{10}$	0.048	0.117	0.275	0.462

Table 7: Empirical power based on $M = 10^3$ independent statistics. Here the significance level is 0.05. The values of the parameter change from $\theta_0 = 1.0$ to $\theta_1 = 1.5$ at time $u_* = \frac{1}{2}, \frac{3}{4}$ or $\frac{9}{10}$.

6 Example C: Cox's regression model

Recall the description of Example C in Section 1.3. Since all the arguments are similar to those in Section 5, we state only the key points in the discussion on the change point problem.

Introducing the notations

$$\begin{aligned}
S_t^{n,0}(\theta) &= \sum_{k=1}^n e^{\theta Z_t^k} Y_t^k, \\
S_t^{n,1}(\theta) &= \sum_{k=1}^n Z_t^k e^{\theta Z_t^k} Y_t^k, \\
S_t^{n,2}(\theta) &= \sum_{k=1}^n (Z_t^k)^\top Z_t^k e^{\theta Z_t^k} Y_t^k,
\end{aligned}$$

we suppose that

$$\sup_{\theta \in \Theta} \sup_{t \in [0, T]} \left\| \frac{1}{n} S_t^{n,l}(\theta) - \mathcal{S}_t^l(\theta) \right\| \rightarrow^p 0, \quad l = 0, 1, 2,$$

where the limits $t \rightsquigarrow \mathcal{S}_t^l$ are some stochastic processes (c.f. Andersen and Gill (1982) who assumed that \mathcal{S}^l 's are not random).

Then, some arguments similar to Section 5.2 are possible for

$$\begin{aligned}
\dot{M}_{\theta_0}(1, \theta) &= \int_0^T \left(\frac{\mathcal{S}_t^1(\theta_0)}{\mathcal{S}_t^0(\theta_0)} - \frac{\mathcal{S}_t^1(\theta)}{\mathcal{S}_t^0(\theta)} \right) \mathcal{S}_t^0(\theta_0) \alpha(t) dt, \\
\dot{M}(u, \theta) &= \int_0^{(u \wedge u_*)^T} \left(\frac{\mathcal{S}_t^1(\theta_0)}{\mathcal{S}_t^0(\theta_0)} - \frac{\mathcal{S}_t^1(\theta)}{\mathcal{S}_t^0(\theta)} \right) \mathcal{S}_t^0(\theta_0) \alpha(t) dt \\
&\quad + 1\{u > u_*\} \int_{u_*^T}^{u^T} \left(\frac{\mathcal{S}_t^1(\theta_1)}{\mathcal{S}_t^0(\theta_1)} - \frac{\mathcal{S}_t^1(\theta)}{\mathcal{S}_t^0(\theta)} \right) \mathcal{S}_t^0(\theta_1) \alpha(t) dt, \\
V_n(u, \theta_0) &= \frac{1}{n} \int_0^{u^T} \frac{\mathcal{S}_t^{n,0}(\theta_0) \mathcal{S}_t^{n,2}(\theta_0) - \mathcal{S}_t^{n,1}(\theta_0) \mathcal{S}_t^{n,1}(\theta_0)^\top}{\mathcal{S}_t^{n,0}(\theta_0)} \alpha(t) dt, \\
V(u, \theta) &= \int_0^{(u \wedge u_*)^T} \frac{\mathcal{S}_t^0(\theta) \mathcal{S}_t^2(\theta) - \mathcal{S}_t^1(\theta) \mathcal{S}_t^1(\theta)^\top}{\mathcal{S}_t^0(\theta)^2} \mathcal{S}_t^0(\theta_0) \alpha(t) dt \\
&\quad + 1\{u > u_*\} \int_{u_*^T}^{u^T} \frac{\mathcal{S}_t^0(\theta) \mathcal{S}_t^2(\theta) - \mathcal{S}_t^1(\theta) \mathcal{S}_t^1(\theta)^\top}{\mathcal{S}_t^0(\theta)^2} \mathcal{S}_t^0(\theta_1) \alpha(t) dt, \\
\hat{V}_n(u) &= \frac{1}{n} \sum_{k=1}^n \int_0^{u^T} \frac{\mathcal{S}_t^{n,0}(\hat{\theta}_n) \mathcal{S}_t^{n,2}(\hat{\theta}_n) - \mathcal{S}_t^{n,1}(\hat{\theta}_n) \mathcal{S}_t^{n,1}(\hat{\theta}_n)^\top}{\mathcal{S}_t^{n,0}(\hat{\theta}_n)^2} dN_t^k.
\end{aligned}$$

Acknowledgements. This work was supported by Prin09 grant (I.N.) and by Grant-in-Aid for Scientific Research (C), 24540152, from Japan Society for the Promotion of Science (Y.N.).

References

- [1] Aït-Sahalia, Y. (1999). Transition densities for interest rate and other nonlinear diffusion. *J. Finance* **54**, 1361-1395.
- [2] Andersen, P.K. and Gill, R.D. (1982). Cox's regression models for counting processes: A large sample study. *Ann. Statist.* **10**, 1100-1120.
- [3] Cox, D.R. (1972). Regression models and life-tables (with discussion). *J. Roy. Statist. Soc. B* **34**, 187-220.
- [4] Fujii, T. and Nishiyama, Y. (2011). Testing for parameter change in a stress release process. *Preprint*.
- [5] Horváth, L. and Parzen, E. (1994). Limit theorems for Fisher-score change processes. In: *Change-point Problems, (Edited by Carlstein, E., Müller H.-G. and Siegmund, D.)* IMS Lecture Notes – Monograph Series **23**, 157-169.
- [6] Iacus, S.M. and Yoshida, N. (2012). Estimation for the change point of volatility in a stochastic differential equation. *Stochastic Process. Appl.* **122**, 1068-1092.

- [7] Ibragimov, I.A. and Has'minskii, R.Z. (1981). *Statistical Estimation: Asymptotic Theory*. Springer-Verlag, New York.
- [8] Isham, V. and Westcott, M. (1979). A self-correcting point processes. *Stochastic Process. Appl.* **8**, 335-347.
- [9] Kato, K. (2011). A note on moment convergence of bootstrap M -estimators. *Statist. Decision* **28**, 51-61.
- [10] Kessler, M. (1995). Estimation des paramètres d'une diffusion par des contrastes corrigés. *C. R. Acad. Sci. Paris Ser. I Math.* **320**, 359-362.
- [11] Kessler, M. (1997). Estimation of an ergodic diffusion from discrete observations. *Scand. J. Statist.* **24**, 211-229.
- [12] Kutoyants, Yu.A. (1984). *Parameter Estimation for Stochastic Processes*. Heldermann, Berlin.
- [13] Kutoyants, Yu.A. (1994). *Identification of Dynamical Systems with Small Noise*. Kluwer Academic Publishers, Dordrecht.
- [14] Kutoyants, Yu.A. (1998). *Statistical Inference for Spatial Poisson Processes*. Lect. Notes Statist. **134**, Springer-Verlag, New York.
- [15] Kutoyants, Yu.A. (2004). *Statistical Inference for Ergodic Diffusion Processes*. Springer-Verlag, London.
- [16] Lee, S., Ha, J., Na, O. and Na, S. (2003). The cusum test for parameter change in time series models. *Scand. J. Statist.* **30**, 781-796.
- [17] Negri, I. and Nishiyama, Y. (2011). Asymptotically distribution free test for parameter change in a diffusion process model. *To appear in Ann. Inst. Statist. Math.*
- [18] Nishiyama, Y. (2010). Moment convergence of M -estimators. *Statist. Neerlandica* **64**, 505-507.
- [19] Nishiyama, Y. (2011). *Statistical Analysis by the Theory of Martingales*. (In Japanese.) ISM Series **1**, Kindaikagakusha, Tokyo.
- [20] Song, J. and Lee, S. (2009). Test for parameter change in discretely observed diffusion processes. *Statist. Inference Stoch. Process.* **12**, 165-183.
- [21] Uchida, M. and Yoshida, N. (2012). Adaptive estimation of an ergodic diffusion process based on sampled data. *Stochastic Process. Appl.* **122**, 2885-2924.
- [22] van der Vaart, A.W. (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge.

- [23] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer-Verlag, New York.
- [24] Yoshida, N. (2011). Polynomial type large deviation inequalities and quasi-likelihood analysis for stochastic differential equations. *Ann. Inst. Statist. Math.* **63**, 431-479.